BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Schillinger *et al.*	Group Art Unit: 1661
Serial No.: 10/052,771	Examiner: Para, A.
Filed: January 23, 2002	Atty. Dkt. No.: ASGR:002USD1
For: PLANTS HAVING RESISTANCE TO MULTIPLE HERBICIDES AND ITS USE	

CERTIFICATE OF MAILING
37 C.F.R. §1.8

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, DC 20231, on the date below:

05/25/04
Date

Robert E. Hanson

DECLARATION OF DR. JOSEPH R. BYRUM UNDER 37 C.F.R. § 1.132

Assistant Commissioner for Patents
Washington, D.C. 20231

I, JOSEPH R. BYRUM HEREBY DECLARE AS FOLLOWS:

1. I am a U.S. citizen and currently reside at 913 31st Street, West Des Moines, Iowa.

2. I have been employed by Asgrow Seed Company, Inc., and Monsanto Company, the parent company of Asgrow Seed Company, Inc., since 1995; currently with the title of Director, Soybean Trait Integration.

3. I hold a Ph.D. in Plant Breeding and Genetics from Iowa State University and a B.S. in Crop and Soil Sciences from Michigan State University. I have been conducting research in the area of plant biochemistry, molecular biology and genetics since 1989. My duties have included the

creation of transgenic crop varieties since 1993.

4. I am an inventor of the above-captioned patent application and am familiar with the contents of the patent application.

5. I understand that the Patent and Trademark Office Examiner in charge of assessing the patentability of the referenced patent application has rejected the claims as being obvious. In particular, it is my understanding that it has been asserted that it would have been obvious to engineer resistance to the herbicides glyphosate and glufosinate in a single soybean plant.

6. I am therefore providing the present Declaration to submit information demonstrating that it would not have been obvious to those of ordinary skill in the field of agriculture and biotechnology at the time the application was filed to produce a single soybean plant exhibiting resistance to glyphosate and glufosinate.

7. As indicated in the application, a soybean variety had not been developed having more than one herbicide resistance trait prior to the current invention. Any prior assertions that glyphosate and glufosinate herbicide resistance transgenes could be successfully expressed in a single variety would therefore be speculation.

The expression of a herbicide resistance transgene requires manipulation of complex metabolic pathways of plant cells. Herbicides such as glyphosate and glufosinate normally interfere with these pathways. For example, glyphosate inhibits the enzyme 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) and glufosinate the phosphinothricin acetyl transferase (PAT) enzyme (see Scheme 1 on page 13 of the patent application). These enzymes are both involved in the synthesis of amino acids in plant cells. The lack of function of the

enzymes kills the plant. Engineering resistance to the herbicides requires modifying and/or overexpression of altered forms of the enzymes with decreased herbicide susceptibility.

Because soybean plants do not naturally exhibit herbicide tolerance, the results of a given modification are unpredictable. Transgene expression causes complex pleiotropic effects that may or may not be detectable at the phenotypic level. The effects can vary depending upon factors such as the location of insertion of the transgene in a genome, the transgene being expressed, the genotype of the host soybean plant, and the regulatory elements and any enhancers used to express the transgene. The expression of enzymes not normally present in the plants also creates a "metabolic drag" reducing energy from the diversion of resources to the expression of the transgene. Herbicide resistance traits add the uncertainty of potential interactions among interrelated metabolic pathways, including negative or positive feedback regulation of different pathways from altered substrate or precursor production.

The above difficulties are not merely additive when combining more than one herbicide resistance transgene. Exponential increases in difficulty arise from the interaction of different traits. First, particular combinations of herbicides have a synergistic effect. Tolerance to glyphosate and glufosinate individually would therefore not necessarily be expected to be indicative of tolerance to the combination of these herbicides. Second, metabolic drag and/or pleiotropic effects could have limited the availability of substrates in the metabolic pathways necessary for the co-expression of glyphosate and glufosinate tolerance transgenes. Third, certain traits are known to be negatively correlated. For example, with a few exceptions, increases in protein in soybeans have resulted in decreased oil based on negative genetic correlations (see, *e.g.*, p. 1038, Wilcox and Cavins. 1995. *Crop Sci.* 35:1036-1041; see also p.3,

Openshaw and Hadley. 1984. *Crop Sci.* 24:1-4). Negative correlations have also been observed between oil content and yield and protein (see, *e.g.*, p. 126 of Hanson *et al.*, *Crop Sci.* 1: 121-126). Absent evidence to the contrary, a negative correlation could also have been observed for the combination of glyphosate and glufosinate tolerance traits.

As a soybean plant had never previously been engineered expressing any two herbicide resistance traits, success with combining glyphosate and glufosinate tolerance simply could not have been reasonably expected based on the prior knowledge in the art. Any one or more of the difficulties mentioned above could have prevented the co-expression of transgenes conferring glyphosate and glufosinate tolerance. Given the multitude of variables, the absence of pleiotropic or other effects preventing the combination of these traits would have been speculation prior to the studies presented in the application.

8. Based on the foregoing, one of skill in the art would have been without any reasonable expectation at the time the application was filed that introducing glyphosate and glufosinate resistance transgenes into a single soybean plant would yield a plant that is resistant to both herbicides.

9. I hereby declare that all statements made herein of my knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

5/24/2004
Date

Joseph R. Byrum

Prout, T. 1962. The error variance of the heritability estimate obtained from selection response. Biometrics 18:404–407.
Rachie, K., and J.V. Majumdar. 1980. Pearl millet. Pennsylvania State Univ. Press, University Park, PA.
Russell, R.S. 1977. Plant root systems: Their functions and interaction with the soil. McGraw-Hill Book Co., London.
SAS. 1988. SAS/STAT. User'sguide. Release 6.03 (ed.) SAS Institute, Inc., Cary, NC.
Soman, P., K.N. Rai, and F.R. Bidinger. 1989. Dwarfing gene effects on coleoptile length in pearl millet. Crop Sci. 29:956–958.
Sunderman, D.W. 1964. Seedling emergence of winter wheats and its association with depth of sowing, coleoptile length under various conditions, and plant height. Agron. J. 56:23–25.
Takahashi, N. 1978. Adaptive importance of mesocotyl and coleoptile growth in rice under different moisture regimes. Aust. J. Plant Physiol. 5:511–517.
Townley-Smith, T.F., and E.A. Hurd. 1979. Testing and selection for drought resistance in wheat. p. 447–464. *In* H. Mussel and R.C. Staples (ed.) Stress physiology in crop plants. John Wiley and Sons, New York.

Backcrossing High Seed Protein to a Soybean Cultivar

James R. Wilcox* and James F. Cavins

ABSTRACT

An inverse relationship between seed yield and seed protein concentration has limited success in developing soybean [*Glycine max* (L.) Merr.] cultivars with high seed protein. High protein from the donor parent 'Pando' (498 g kg^{-1} protein) was backcrossed to 'Cutler 71' (408 g kg^{-1} protein) to determine if the yield of Cutler 71 could be recovered in addition to the high protein from Pando. Random F_4-derived lines, plus three lines with highest seed protein concentration, from the initial cross, the BC_1, and the BC_2 populations, were evaluated for agronomic traits in separate, two-replicate tests for 1 yr at West Lafayette, IN. Seed from replication composites were evaluated for protein and oil concentration using near infra-red reflectance or near infra-red transmission. The parent line for each backcross was selected first for high seed protein, then for yield and agronomic similarity to Cutler 71. Random F_4-derived progenies of the BC_3 population, the parent line for each backcross, and the cultivars Pando, Cutler 71, and Hamilton were evaluated in three-replicate tests for 2 yr at West Lafayette, IN. In each backcross generation, lines were identified with seed protein in excess of 470 g kg^{-1} and that progressively approached the yield of Cutler 71. In the BC_3 population, one line averaged 472 g kg^{-1} seed protein and was significantly ($P = 0.05$) higher in seed yield than Cutler 71, similar in yield to the cultivar Hamilton. In each backcross population, there were inverse relationships between seed yield and seed protein (R^2 values ranging from 0.33 to 0.06) and between seed protein and seed oil (R^2 values ranging from 0.55 in BC_1 to 0.94 in BC_3). In successive backcross populations, minimum oil values increased from 148 in BC_1 to 174 g kg^{-1} in BC_3, indicating a trend toward recovering oil concentration (204 g kg^{-1}) of Cutler 71. The data demonstrate that high seed protein can be backcrossed to a soybean cultivar, fully recovering the seed yield of the cultivar, suggesting the absence of physiological barriers to combining high seed protein with high seed yield in these populations.

Success in breeding soybean cultivars with higher-than-normal seed protein has been hampered by an inverse relationship between seed yield and seed protein concentration. Burton (1984), summarizing results of several breeding studies, reported genotypic correlations between seed yield and seed protein percentage varied from −0.12 to −0.74. In only one population was there a positive genotypic correlation between these two traits. Additional studies by Sebern and Lambert (1984), Simpson and Wilcox (1983), and Wehrmann et al. (1987) reported moderate to strong inverse relationships between seed yield and seed protein with correlation coefficients ranging from −0.23 to −0.86.

Previous attempts to backcross high seed protein to high yielding soybean cultivars and breeding lines have not been successful. Hartwig and Hinson (1972) evaluated BC_1 and BC_2 progenies from crosses between the high yielding recurrent parent, 'Bragg', and a high protein parent averaging 112% the protein content of Bragg. Selection for seed oil, based on nuclear magnetic resonance, was effective in obtaining lines that differed in seed protein because of the negative relationship between seed oil and protein. The highest yielding BC_2 lines that were similar to Bragg in seed yield and were high, intermediate, and low in oil averaged only 397, 418, and 435 g kg^{-1} protein, respectively.

Cianzio and Fehr (1982) evaluated seed protein and oil of F_2-derived lines in the F_2 generation and BC_1F_1-derived and BC_2F_1-derived lines in the F_3 generation of crosses between the high protein lines Pando and PI 153.269 and the high yielding cultivars Wells and Woodworth. No line from either set of crosses had protein concentrations as high as those of the high protein donor parent. Mean protein percentages and genetic variances of the populations decreased with each backcross to the high yielding parent. Their results indicated to them that it will be difficult to transfer genes for extremely high protein levels to cultivars with lower protein, by backcrossing to the low protein parent. No yield data were recorded on the breeding lines evaluated in this study.

Wehrmann et al. (1987) evaluated 95 BC_2 progenies in each of three populations, where the recurrent parents were high yielding lines and the donor parent was Pando, that averaged 480 g kg^{-1} seed protein. In these populations, no backcross-derived lines were recovered that combined exceptionally high seed protein with the yield of the recurrent parent. In each of two populations, the highest protein line averaged only 422 and 433 g kg^{-1} protein and did not differ significantly in yield or seed oil from the recurrent parent. In the third population, the highest protein line averaged 462 g kg^{-1} protein

J.R. Wilcox, USDA–ARS, Crop Production and Pathology Res. and Dep. of Agronomy, Purdue Univ., W. Lafayette, IN 47907; and J.F. Cavins, retired, USDA–ARS, Analytical Chemistry, Natl. Center for Agric. Utilization Res., 1815 N. University St., Peoria, IL 61604. Journal Paper no. 14341 of the Purdue Univ. Agric. Res. Programs. Received 11 Aug. 1994. *Corresponding author (jwilcox@dpet.agry.purdue.edu).

Published in Crop Sci. 35:1036–1041 (1995).

but was significantly lower in both yield and seed oil concentration than the recurrent parent.

None of the backcross studies evaluated progenies beyond the BC_2 generation. The lack of success in transferring exceptionally high seed protein to high yielding cultivars by backcrossing casts doubt on the possibility of combining these two traits in adapted germplasm or cultivars.

Our study was initiated to determine if, using successive backcrosses, the high seed protein concentration of Pando could be successfully transferred into a commercial soybean cultivar, fully recovering the yield of the cultivar. This would demonstrate the absence of genetic or physiological barriers to breeding high yielding, high protein cultivars. The germplasm developed would also be useful in breeding high protein cultivars competitive in yield with existing cultivars.

MATERIALS AND METHODS

The Maturity Group IV cultivar Cutler 71 (Probst et al., 1971), averaging 410 g kg^{-1} protein and 210 g kg^{-1} oil on a moisture-free basis, was selected as the recurrent parent. When the initial cross was made, Cutler 71 was a high yielding Maturity Group IV cultivar adapted to production in central and southern Indiana (Wilcox et al., 1976). The Maturity Group 00 accession Pando with 522 g kg^{-1} protein and 155 g kg^{-1} oil was used as the high protein donor parent (Bernard et al., 1995). Cutler 71 was crossed with Pando in 1971 and the F_2 through F_4 generations advanced by single seed descent. Two-hundred-and-three F_4 plants were selected and grown in F_5 progeny rows 2 m long and spaced 0.75 m apart. Data were recorded on plant maturity and height, individual rows were harvested, and protein and oil were determined on a 25-g sample of harvested seed. A random sample of 45 F_4-derived lines and three lines with the highest seed protein were evaluated in a two-replicate yield test, planted 12 May 1976. Four-row plots 5 m long with 0.75 m between rows were planted with 450 seeds. Plots were end trimmed to 4 m prior to harvest. Data were recorded on the center two rows of each plot for the characteristics maturity (95% pods brown), plant height (centimeters from soil surface to tip of plants), and lodging (scored from 1, all plants erect, to 5, all plants prostrate). The center two rows of each plot were harvested and seed yield recorded in grams per plot. A 25-g sample of seed, composited across replications, was analyzed for protein and oil concentration on a moisture-free basis. A line was selected first, based on high seed protein, then on similarity to the recurrent parent for agronomic traits measured. This line was used as the parent for the backcross to Cutler 71.

The first backcross was made in 1977 and the F_2 through F_4 generations advanced by single seed descent. One-hundred-sixty-six single F_4 plants were selected and grown in 2-m F_5 progeny rows, spaced 0.75 m apart. Data were recorded as on the progeny rows from the initial cross. The F_4-derived lines from the BC_1 population were evaluated in three tests, one consisting of early maturing lines, the second of midseason maturing lines, and the third of late maturing lines. Data on 38 of the F_4-derived lines of midseason maturity, 35 of which represented random selections and three of which were high in seed protein concentration, are reported here. The 38 lines were evaluated in a two-replicate yield test planted 23 May 1982. Plot size and shape were identical to that used for populations from the initial cross and the same data were recorded. As in the initial cross, a line was selected first on the basis of high seed protein, then on similarity in seed yield and agronomic traits to the recurrent parent. This line was used as the parent for the BC_2 cross.

The BC_2 cross was made in 1983 and the F_2 through F_4 generations advanced by single seed descent. Two-hundred-forty-two F_4 plants were selected and grown in F_5 plant rows 2 m long and spaced 0.75 m apart. Data were recorded as on plant rows in the original and the BC_1 cross. A random sample of 73 BC_2 lines and three lines with the highest seed protein were evaluated in a two-replicate yield test planted 17 May 1988. Plot size and shape were similar to those used to test progenies from the original and the BC_1 cross except within-plot row spacing was 0.60 m. Data were recorded as on replicated test plots of the original and BC_1 progenies. As in previous generations, a line was selected first on the basis of high seed protein, then on similarity of yield and agronomic traits to Cutler 71. This line was used as the parent for the BC_3 cross.

The BC_3 cross was made in the greenhouse during the winter, 1987 to 1988. The F_2 through F_4 generations were advanced by single seed descent. Two-hundred-thirty-seven F_4 plants were selected and grown in F_5 plant rows 2 m long. Data were recorded on plant rows as in the previous populations. A random sample of 71 BC_3 lines and the three lines with the highest seed protein were selected for evaluation in replicated yield tests. Also included in the yield tests were the donor and the recurrent parents, the lines selected as the parent for each backcross, and Hamilton (Nickell et al., 1990), a currently grown cultivar similar in maturity to Cutler 71 and the majority of the BC_3 lines. Hamilton was included as a currently grown yield comparison for the BC_3 lines and the older cultivar, Cutler 71. The BC_3 lines and check entries were evaluated in three-replicate yield tests in 1991 and 1992. Plot size and shape were identical to that used for the BC_2

Table 1. Performance of the selected parent from each backcross generation compared with that of the donor parent, the recurrent parent, and Hamilton, averaged across three replications and 2 yr.

Strain	Parent or selection	Seed			Mature date	Plant height	Lodging‡
		Yield	Protein†	Oil†			
		kg ha^{-1}	g kg^{-1}			cm	
Pando	Donor	800	498	148	6 Aug.	28	1.0
CX602-107	BC_1 parent	2145	474	175	17 Sept.	109	1.8
CX797-21	BC_2 parent	2179	476	167	25 Sept.	109	1.4
CX1038-14	BC_3 parent	2529	466	171	3 Oct.	117	1.9
CX1307-205	BC_3 selected line	2831	472	174	28 Sept.	117	1.6
Cutler 71	Recurrent	2461	408	204	27 Sept.	109	1.8
Hamilton	Control	2885	402	214	25 Sept.	91	1.3
LSD 0.05		200	7	5	4	5	0.3

† Moisture-free basis.
‡ Score: 1 (all plants erect) to 5 (all plants prostrate).

progenies. Data were recorded on each plot as in the tests of progenies from the original cross and the BC_1 and BC_2 progenies. Protein and oil concentrations of the seed were determined on a 25-g sample from each replication in each year of these tests.

The F4 generation, the F_5 plant rows, and the replicated yield tests of each population were grown on a Chalmers soil (fine-silty, mixed, mesic Typic Haplaquoll) at the Purdue Univ. Agron. Res. Center near West Lafayette, IN. Protein and oil concentrations were determined at the USDA–ARS National Center for Agricultural Utilization Research, Peoria, IL, by near infra-red reflectance for the initial and BC_1 populations and by near infra-red transmission for the BC_2 and BC_3 populations. Analyses of variance for randomized complete-block designs were used to evaluate the data from the replicated yield tests of progenies from the original, BC_1, BC_2, and BC_3 populations. The F_4-derived lines were considered random effects in each of these analyses. Years were considered random effects in the 2-yr analyses of the F_4-derived lines from the BC_3 population. Regression analyses were based on progeny means for each of the above populations.

Hanson's (Leffel, 1990) approximate processed value was calculated for each parental line, the high-protein BC_3 selection, and Hamilton based on 2-yr mean performance data of the BC_3 population. Oil and protein prices were the average prices for the 12-mo period January through December, 1993 (USDA, 1994).

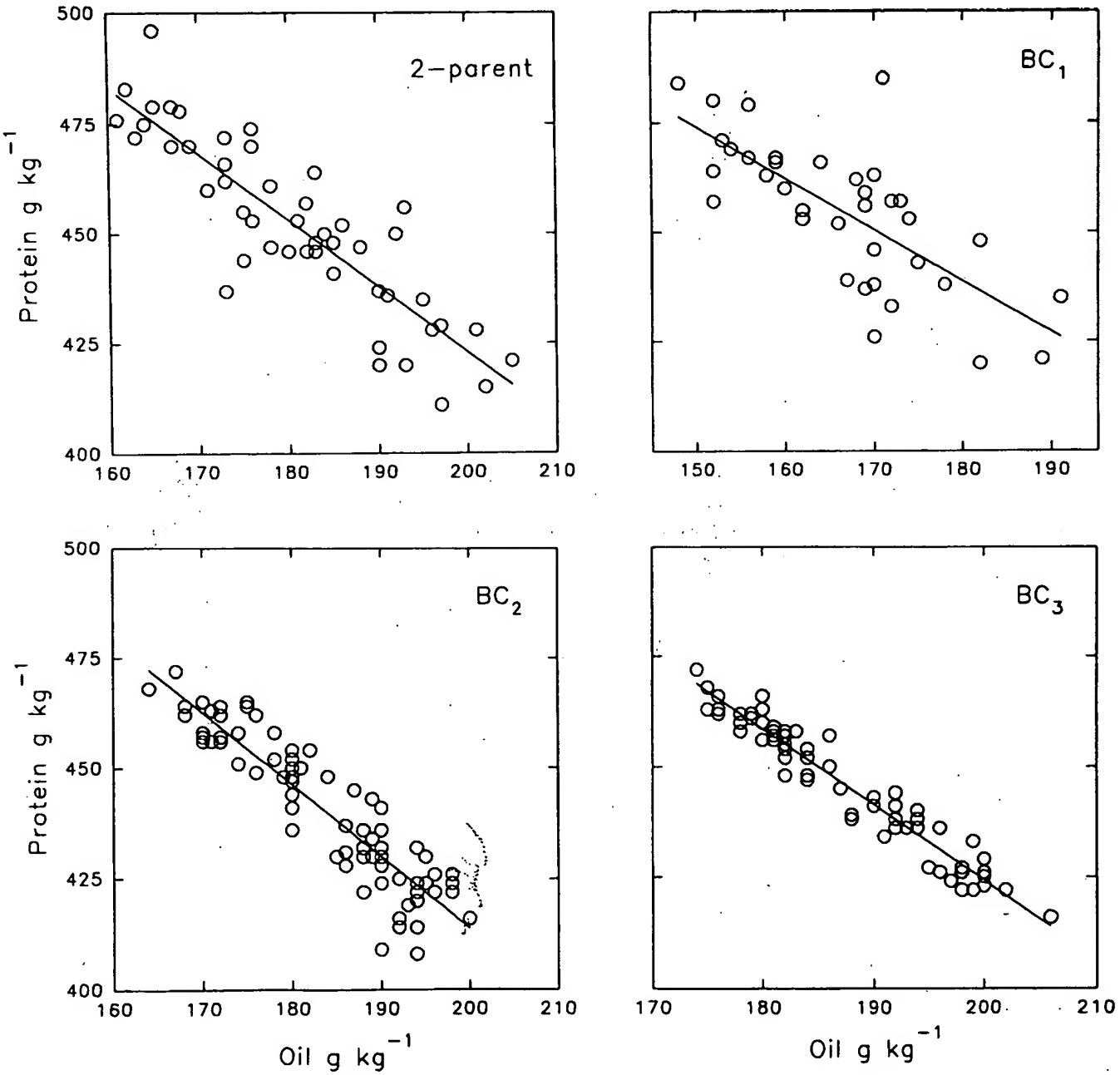

Fig. 1. Regression of seed protein concentration on seed oil concentration in successive backcross populations of Cutler 71 (recurrent parent) × Pando (donor parent). Note changes in scale of abscissa. Regression values in Table 2.

RESULTS AND DISCUSSION

Data obtained in the 2-yr test that included the donor parent, Pando, and the recurrent parent, Cutler 71, were consistent with previous reports on the performance of these cultivars. Pando averaged 800 kg ha^{-1} seed yield, matured 6 Aug., and averaged 498 g kg^{-1} protein and 148 g kg^{-1} oil (Table 1). In contrast, Cutler 71 averaged 2461 kg ha^{-1} seed yield, matured 27 Sept., and averaged 408 g kg^{-1} protein and 204 g kg^{-1} oil.

There was a wide range in seed protein and oil concentrations among F_4-derived lines in the initial two-parent cross and in each of the backcross populations as evaluated in their respective tests (Fig. 1). This variability provided opportunities to select a line in each population with high seed protein and, in successive backcross populations, agronomic characteristics progressively more similar to those of Cutler 71.

With successive backcrosses, rapid progress was made in recovering the yield and other agronomic traits of Cutler 71 while maintaining high seed protein from the donor parent, Pando (Table 1). The F_4-derived lines selected as parents from the two-parent cross, CX602-107, and from the BC_1 population, CX797-21, were similar in seed yield and seed protein concentration. Yields of both lines

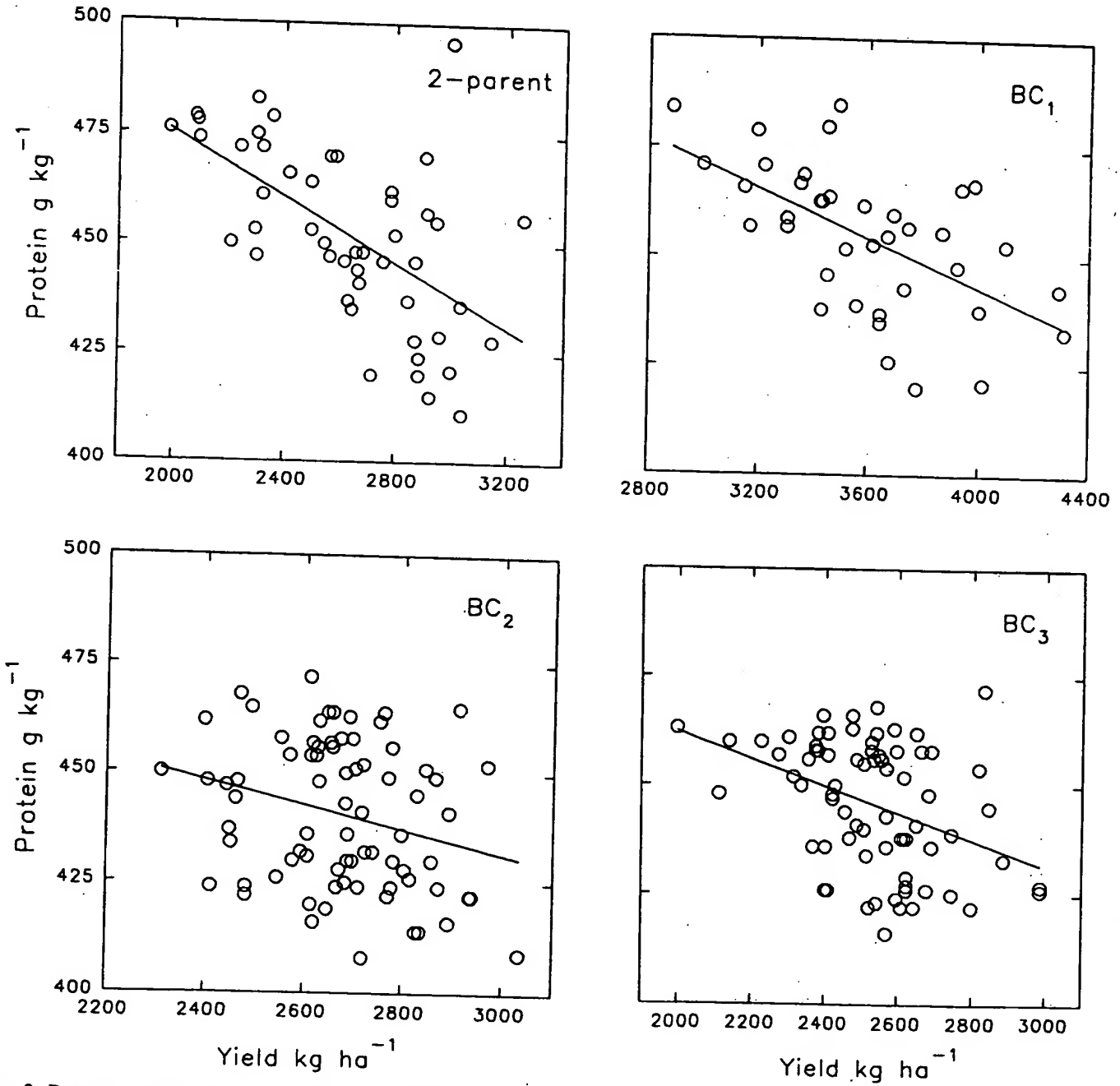

Fig. 2. Regression of seed protein concentration on seed yield in successive backcross populations of Cutler 71 (recurrent parent) × Pando (donor parent). Note changes in scale of abscissa. Regression values in Table 2.

were well above the yield of Pando and approached the yield of Cutler 71. Among progenies of the BC_2 population, the line CX1038-14 was similar in yield but matured 5 d later than Cutler 71. The F_4-derived BC_3 line CX1307-205 averaged significantly ($P = 0.05$) higher in seed yield than Cutler 71, similar in yield to the cultivar Hamilton, and averaged 472 g kg^{-1} seed protein in these tests. Backcross-derived soybean lines that exceed the yield of the recurrent parent have been reported in BC_3 and BC_4 populations and even in BC_7 populations (Wilcox et al., 1971). The selected BC_3 line CX1307-205 was similar in maturity date and lodging score to Cutler 71 but averaged 8 cm taller in mature plant height than the recurrent parent.

Hanson et al. (1961) determined energy requirements per unit area needed to produce protein, oil, and protein plus oil. From their analysis, they concluded there should be no restrictions in producing a high yielding, low oil soybean and, if N is not limiting or other physiological restrictions are not present, a high yielding, high protein soybean. The performance of CX1307-205 demonstrates that high yielding, high protein soybean cultivars can be a reality and there does not appear to be any physiological barrier to their development.

Results suggest that not all of the genes controlling seed protein concentration in Pando were transferred to the recurrent parent. The line with the highest seed protein from the two-parent cross was not equal to the 498 g kg^{-1} protein concentration in the seed of Pando. In successive backcross populations, the high-protein selection was significantly lower in seed protein than Pando. However, backcrossing was successful in transferring seed protein concentrations as high as 472 g kg^{-1} from Pando to BC_3 progenies that were equal to or greater in seed yield than Cutler 71.

Wehrmann et al. (1987) also used Pando as a source of high seed protein in attempts to transfer this trait to three high yielding soybean cultivars. They were successful in recovering the high yields of the recurrent parents after two backcrosses but only moderately successful in increasing seed protein concentration. The highest yielding BC_2F_2-derived lines averaged 373, 376, and 427 g kg^{-1} protein compared with 361, 362, and 413 g kg^{-1} for their respective recurrent parents. The highest protein line in two populations averaged 422 and 433 g kg^{-1} seed protein, considerably below the 480 g kg^{-1} of the donor parent, Pando. In the third population, the highest protein line averaged 462 g kg^{-1} protein but was significantly lower in yield than the recurrent parent.

Success in combining 472 g kg^{-1} protein with the high yield of the recurrent parent in these populations may be due in part to delaying selection for seed protein concentration until the F_4 generation when plants should be fairly homozygous for loci controlling this trait. Wehrmann et al. (1987) based their selections for seed protein on evaluation of seed produced on F_2 and F_3 plants. Heterozygosity at loci controlling seed protein concentration in these early generations may have limited effectiveness of identifying parents that would consistently transmit loci controlling high protein concentrations to progenies of successive backcrosses.

Table 2. Regression values for protein concentration (dependent variable) on yield and oil concentration (independent variables) for progenies from the two-parent cross Cutler 71 × Pando and successive backcrosses to Cutler 71.

Progeny	No. of lines	Mean squares due to		Slope (b)	SE† (b)	R^2
		Regression	Error			
		Protein (g kg^{-1}) vs. yield (kg ha^{-1})				
Two-parent	48	6 328	274	−0.037	0.0077	0.33
BC_1	38	3 315	191	−0.028	0.0068	0.32
BC_2	76	1 308	263	−0.028	0.0125	0.06
BC_3	74	2 281	186	−0.031	0.0088	0.15
		Protein (g kg^{-1}) vs. oil (g kg^{-1})				
Two-parent	48	14 372	99	−1.505	0.1250	0.76
BC_1	38	5 645	126	−1.173	0.1755	0.55
BC_2	76	17 330	47	−1.622	0.0840	0.83
BC_3	74	14 709	13	−1.717	0.0517	0.94

† SE = standard error.

The relationships between seed yield and seed protein concentration in successive backcross populations are illustrated in Fig. 2. The slope of the regression line for each population was significantly ($P = 0.05$) different from 0, and there were no differences in slope among regression lines for the different populations (Table 2). This demonstrates a consistent inverse relationship between seed yield and seed protein, even among progenies from the third backcross to Cutler 71. The low R^2 values for the BC_2 and BC_3 progenies suggest there were greater opportunities to identify high yielding, high protein lines among these progenies than among the two-parent and BC_1 progenies.

There was a strong inverse relationship between seed protein and seed oil among progenies of each of the populations evaluated (Fig. 1). This is consistent with reports of Burton (1984), Hartwig and Hinson (1972), Sebern and Lambert (1984), Simpson and Wilcox (1983), and Wehrmann et al. (1987). Slopes of the regression lines when seed protein was regressed on seed oil were all negative, significantly ($P = 0.05$) different from 0, and similar for the two-parent, BC_2, and BC_3 populations (Table 2). Averaged across the four populations, there was a 1.5 unit decrease in seed protein per unit increase in seed oil. This is similar to the 1.5 to 1.6% gain in seed protein for each 1% decrease in seed oil calculated by Hanson et al. (1961) in a genetic analysis of energy production in the soybean.

Deviations from regression of protein concentration on oil concentration were progressively smaller and R^2 values were progressively larger with successive backcrosses to Cutler 71 (Table 2). This indicates a stronger inverse relationship between seed protein and seed oil among progenies of successive backcrosses to the recurrent parent. However, minimum values for seed oil increased among progenies of the BC_1, BC_2, and BC_3 populations (Fig. 1). This indicates that not only was the yield of the recurrent parent recovered but there was a tendency toward recovering the oil concentration of the recurrent parent as well. Maximum oil values would not be expected to increase in successive backcross populations above $\approx$204 g kg^{-1}, the average value for Cutler 71, the recurrent parent.

Table 3. Approximate processed value based on chemical composition of seed of soybean parental selections and Hamilton. Approximate processed value based on data averaged across three replications in each of 2 yr, an oil price of \$0.504 kg^{-1}, and a protein meal price of \$0.467 kg^{-1}.

Strain	Oil kg^{-1} seed	Protein kg^{-1} seed	Total kg^{-1} seed	Value ha^{-1}
		\$		
Pando	0.0650	0.2026	0.2676	214.08
CX602-107	0.0767	0.1927	0.2694	577.83
CX797-21	0.0732	0.1936	0.2668	581.36
CX1038-14	0.0750	0.1894	0.2644	667.71
CX1307-205	0.0763	0.1918	0.2681	758.99
Cutler 71	0.0895	0.1660	0.2555	628.79
Hamilton	0.0939	0.1634	0.2573	742.31

The economic values of high protein strains compared with the values of high yielding strains with high seed oil concentration were compared. Approximate processed values, which are functions of seed protein and oil concentration and the price of each component, were calculated for parent strains, the high-protein BC_3 selection, and Hamilton (Table 3). Cutler 71 and Hamilton, with oil concentrations typical of currently grown cultivars, had the highest monetary values for seed oil of the lines evaluated. Pando, the high protein donor parent, had the highest monetary value for seed protein. Seed protein values for each of the backcross parents and CX1307-205 were higher than protein values for Cutler 71 and Hamilton. This resulted in higher total approximate processed value for the high protein parent and selections than for these two cultivars. When expressed on a hectare basis, CX1307-205 had the highest total approximate processed value, higher than that of the recurrent parent or the cultivar Hamilton. This demonstrates that soybean strains with high seed protein can be developed that are as profitable, on a chemical constituent basis, as currently grown cultivars.

This research demonstrates that seed yield of a recurrent parent can be recovered when backcrossing high seed protein to the recurrent parent. Selection for the recurrent phenotype as a secondary trait to protein in each backcross population was effective in the recovery of the Cutler 71 phenotype with only three backcrosses. An unexpected benefit was the trend toward recovering oil of the recurrent parent while maintaining high seed protein from the donor parent. The approximate processed value calculations demonstrated that the highest yielding, high protein BC_3 line had a higher chemical constituent value than a currently grown cultivar of similar maturity.

ACKNOWLEDGMENT

The authors acknowledge the support of the Indiana Soybean Development Council, who provided part of the funding for this research.

REFERENCES

Bernard, R.L., C.R. Cremeens, R.L. Cooper, F.I. Collins, O.A. Krober, K.L. Athow, F.A. Laviolette, C.J. Coble, and R.L. Nelson. 1995. Evaluation of the USDA soybean germplasm collection: Maturity groups 000 to IV (FC 01.547 to PI 266.807). USDA Tech. Bull. (in press).

Burton, J.W. 1984. Breeding soybeans for improved protein quantity and quality. p. 361–367. *In* R. Shibles (ed.) Proc. of the World Soybean Res. Conf. III. Ames, IA. 12–17 Aug. 1984. Westview Press, Inc., Boulder, CO.

Cianzio, S.R., and W.R. Fehr. 1982. Genetic variability for soybean seed composition in crosses between high- and low-protein parents. J. Agric. Univ. P.R. 66:123–129.

Hanson, W.D., R.C. Leffel, and R.W. Howell. 1961. Genetic analysis of energy production in the soybean. Crop. Sci. 1:121–126.

Hartwig, E.E., and K. Hinson. 1972. Association between chemical composition of seed and seed yield of soybeans. Crop Sci. 12: 829–830.

Leffel, R.C. 1990. Economic models and breeding strategies for soybean improvement. J. Prod. Agric. 3:582–586.

Nickell, C.D., D.J. Thomas, and P. Stephens. 1990. Registration of 'Hamilton' soybean. Crop Sci. 30:1364.

Probst, A.H., F.A. Laviolette, J.R. Wilcox, K.L. Athow, and T.S. Abney. 1971. Registration of Cutler 71 soybean. Crop Sci. 11: 941–942.

Sebern, N.A., and J.W. Lambert. 1984. Effect of stratification for percent protein in two soybean populations. Crop Sci. 24:225–228.

Simpson, A.M., Jr., and J.R. Wilcox. 1983. Genetic and phenotypic associations of agronomic characteristics in four high protein soybean populations. Crop Sci. 23:1077–1081.

USDA. 1994. Oilseeds: World markets and trade. Circ. Ser. FOP 6-94. Foreign Agricultural Service, Washington, DC.

Wehrmann, V.K., W.R. Fehr, S.R. Cianzio, and J.F. Cavins. 1987. Transfer of high seed protein to high-yielding soybean cultivars. Crop Sci. 27:927–931.

Wilcox, J.R., K.L. Athow, R.J. Martin, F.A. Laviolette, T.S. Abney, T.L. Richards, and M.L. Swearingin. 1976. Performance of soybean varieties in Indiana, 1970–1975. Purdue Univ. Agric. Exp. Stn. Bull. 116., West Lafayette, IN.

Wilcox, J.R., A.H. Probst, K.L. Athow, and F.A. Laviolette. 1971. Recovery of the recurrent parent phenotype during backcrossing in soybeans. Crop Sci. 11:502–507.

ing blocks in a 2-year period, 1956–1957, and progenies were grown from groups of clones selected for superiority of several plant characters in each type. These progenies were evaluated for the same characters as their parents in a 2-year period, 1957–1958.

Interaction of genotype with environment was large and necessarily was removed from estimates of genetic variance. Heritability estimates from variance components and parent-progeny regressions indicate that there is ample additive genetic variance present among clones of switchgrass to make additional progress by selection of clones within superior strains of the grass. This viewpoint is further supported by realized heritabilities and the close agreement of predicted and observed gains for several characters.

Genotypic and phenotypic correlations among the observed characters indicate that effective simultaneous selection may be made for important characters in the development of superior varieties provided that sufficiently large populations are screened.

LITERATURE CITED

1. BURTON, G. W., and DEVANE, E. H. Estimating heritability in tall fescue *(Festuca arundinacea)* from replicated clonal material. Agron. J. 45:478–481. 1953.
2. EBERHART, S. A., and NEWELL, L. C. Variation in domestic collections of switchgrass, *Panicum virgatum* L. Agron. J. 51: 613–616. 1959.
3. FREY, K. J., and HORNER, T. Comparison of actual and predicted gains in barley selection experiments. Agron. J. 47:186–188. 1955.
4. KEMPTHORNE, O. An Introduction to Genetic Statistics. John Wiley & Sons, Inc. 1957.
5. KNEEBONE, W. R. Heritabilities in sand bluestem (*Andropogon hallii* Hack.) as estimated from parental clones and their open-pollination progenies. Agron. J. 50:459–461. 1958.
6. MACMILLAN, CALVIN, and WEILER, JOHN. Cytogeography of *Panicum virgatum* in Central North America. Am. J. Bot. 46: 590–593. 1959.
7. NEWELL, L. C., and TYSDAL, H. M. Numbering and note-taking systems for use in the improvement of forage crops. J. Am. Soc. Agron. 37:736–749. 1945.
8. NIELSEN, E. L. Analysis of variation in *Panicum virgatum*. J. Agr. Res. 69:327–353. 1944.
9. ROBINSON, H. F., COMSTOCK, R. E., and HARVEY, P. H. Estimates of heritability and the degree of dominance in corn. Agron. J. 41:353–359. 1949.

Genetic Analysis of Energy Production in the Soybean[1]

W. D. Hanson, R. C. Leffel, and Robert W. Howell[2]

THE emphasis in soybean breeding has been on the production of high yielding, high oil lines; however, a hift in emphasis from oil to protein has been made to obtain a balanced total program. In practice, an increase in oil percentage has been associated with a decrease in protein percentage. The change in percent oil to the change in ercent protein is considered to be approximately 1 to 2. The basis for this statement or, for that matter, the basic neasures for studying the genetic control of the energy are ot clearly defined. The objective of this paper is to quanify the genetics involving the production of energy by the oybean plant and to relate this information with the resent breeding program.

A plant produces energy by first synthesizing sugar. The nergy which is not used for plant metabolism and growth s stored. The form in which the energy is stored and the uality and the quantity of the stored energy is the unique xpression of a genotype grown within an environment. ield components such as seed size, seeds per fruiting unit, nd units per plant are genetically controlled but must be onsidered subservient to the ability of a genotype to proce energy unless storage limitations exist. Two genetic ystems should be considered: (i) the genetic potential reuired to produce total seed energy and (ii) the genetic ontrol of the distribution of energy between seed fractions.

MATERIALS AND METHODS

Experimental material—The data used for the genetic analysis ere obtained from Leffel et al. (10). For one test environment, 20 F_3 soybean lines were randomly selected from each of 45 F_3 families resulting from a diallel cross system. The procedure was repeated so that four environments were sampled. Thus, 80 F_3 lines were grown from each cross. The environments will be represented as 15, 25, 16 and 26 in which the digits 1 and 2 represent the Maryland locations, Beltsville and Cambridge, and 5 and 6 represent the years, 1955 and 1956. The experimental material was blocked into 20 groups and replicated twice. Each group consisted of one random F_3 line from each cross. The analysis for one environment is given in Table 1. The combined analysis is simply a pooled within-environment analysis since different F_3 lines occurred within each environment. Although the genotype by environment interactions are not available, some of the concepts to be developed involve the description of a relationship within an environment and the consistency of the description between the environments. For a complete evaluation the genotype by environment interaction would be helpful.

Energy concepts—The mineral composition of the soybean seed was taken as 5% and assumed not to contribute significantly to total seed energy. Energy sources are identified as protein (X_1%), oil (X_2%), and residual (X_3%). "Protein" is calculated from total nitrogen and therefore includes all nitrogen-containing substances. "Oil" is the petroleum ether-soluble material and includes a small amount of non-triglycerides. The residual portion not accounted for as protein, oil, or ash ($X_3\% = 95 - X_1\% - X_2\%$) may be considered to be mostly carbohydrate including soluble carbohydrates and structural material such as cellulose. This residual fraction has been represented as sucrose with a caloric value of 3.95 Kcals./gram.

The average caloric value for soybean oil was estimated as 9.40 Kcals./gram. An average fatty acid composition for soybean oil was taken as 5% linolenic, 50% linoleic, 25% oleic, and 20% saturated 18-carbon acids (stearic) (Howell, 6). The small amounts of acids of other than 18-carbon lengths were ignored. The average fatty acid ($C_{18}H_{33.4}O_2$) has a heat of combustion of 2634 Kcals./mole. The heats of combustion for oleic and stearic

TABLE 1—Analysis of variance for F_3 line variability within an environment.

Source of variation	Degrees of freedom	Expected mean square
Reps and Groups	39	
Crosses	44	
Crosses x Groups	836	$\sigma^2_{\epsilon_1} + 2\sigma^2_{g_1}$
Error	880	$\sigma^2_{\epsilon_1}$

$\sigma^2_{g_1}$ estimates the F_3 line variability within crosses and within locations and is taken as an estimate of the additive genetic variability.

[1] Joint contribution from Crops Research Division, Agricultural esearch Service, USDA, and the Maryland and Illinois Agricultural Experiment Stations. Publication No. 347 of the U.S. Reonal Soybean Laboratory, Urbana, Ill. Received for publication ine 13, 1960.

[2] Geneticist, Crops Research Division, Beltsville, Md.; Agronoist, Department of Agronomy, University of Maryland, College ark, Md.; and Plant Physiologist, U.S. Regional Soybean Laboratory, Urbana, Ill. The authors acknowledge the contributions of . A. Tabler, Computing Laboratory, Biometrical Services, USDA, eltsville, Md., to the computational aspects of this paper.

TABLE 2—Amino acid composition assumed for soybean protein (Block_and_Weiss,1).

Amino acid	Grams in 100 g. protein	Amino acid	Grams in 100 g. protein
Glycine	3.8	Glutamic acid	18.6
Alanine	4.5	Methionine	1.1
Valine	5.2	Cystine	1.2
Leucine	7.6	Lysine	6.6
Isoleucine	5.8	Arginine	7.0
Proline	4.1	Histidine	2.5
Phenylalanine	4.8	Tryptophane	1.2
Tyrosine	3.2	Serine	5.6
Aspartic acid	8.3	Threonine	3.9

acids were taken from Kharasch (9). The values for linolenic and linoleic were calculated by assuming that each additional double bond reduced energy by an amount equal to the difference between stearic and oleic. Three moles of fatty acids are esterified with one mole of glycerol ($C_6H_8O_3$, 397 Kcals./mole) giving triglycerides with an average formula of $C_{57}H_{101.6}O_6$ and an average heat of combustion of 8299 Kcals./mole or 9.40 Kcals./g. oil. The small energy change due to the formation of the three ester bonds is neglected.

The average caloric value for protein was estimated as 4.57 Kcal./gram. The amino acid composition considered for soybean protein was taken from Block and Weiss (1) and is summarized in Table 2. The heats of combustion for the amino acids were obtained from Kharasch (9). The heats of combustion for amino acids not listed by Kharasch were estimated from the heats of combustion for similar compounds. The energy in the peptide bonds was neglected. This energy is about 3 to 4 Kcal./bond, which is less than 1% of the energy of the amino acid in most cases.

Carbon equivalent—The products of the first energy-yielding sequence of the plant (glycolysis) are three carbon sugar derivatives. These substances may be channeled into alternate energy forms. An input of energy is required in the synthesis of products, such as oil, with energy levels higher than sugar. The input energy includes the additional energy now stored in the product plus energy expended to make the conversion. This expended energy can be expressed as "work" carbons or the initial sugar carbons required to reach the new energy level. The "work" carbon appears as CO_2 and thus is lost to the plant. Carbon equivalent is defined as the gram atoms of sugar carbon required to produce a finished product including both gram atoms of "work" carbon lost in the synthesis and the gram atoms of carbon stored in the seed fraction. This unit of measure expresses a seed fraction in terms of the weight of sugar carbon required to produce the fraction rather than weight of the fraction.

For oil, one "work" carbon disappears for each 2 carbons present in the fatty acid. Since there are 54 fatty acid carbons and 3 glycerol carbons in the oil formula, the ratio of "work" carbons to oil carbons is $27/57 = .474$. The "work" carbons for protein are much less certain than those for oil because of the uncertainty of the synthetic mechanisms. However, using the information summarized by Fruton and Simmonds (4), one can estimate the "work" carbon required to produce the soybean protein as a difference between the total number of carbons in the primary material involved in the amino acid synthesis and the number in the amino acids. The "work" carbon to protein carbon ratio was estimated as .478. It is assumed that no "work" carbon, or only a negligible quantity, is involved in the formation of the residual fraction.

The conversion of oil or protein to carbon equivalent would also require the factor, (carbon weight in fraction)/(formula weight in fraction). For oil, 57 moles of carbon = 684. Ratio carbon weight/formula weight = $684/883 = .775$. Total oil carbon per 100 grams of seed is therefore $.775X_2$. Similar calculation based on the amino acid distribution from Block and Weiss (1) gives a ratio of .532 for protein. If the residual fraction is represented as glucose, a ratio of .400 is obtained for this fraction. The information on energy, carbon ratios, and "work" carbons is summarized in Table 3. The conversions involving energy are included in the table. The weight of sugar carbon (carbon equivalent) required to produce a gram of soybean seed can be calculated directly from the information given. An interesting item is that about one-fourth of the initial sugar carbons associated with the seed complex is lost as "work" carbon to produce the protein and oil in the soybean seed.

Relative genetic sensitivity—The transformation, energy per acre or the proportion of sugar carbons, requires the combination of two or more measurement criteria. One needs to define a statistic

TABLE 3—Caloric and gram carbon conversion factors for the three soybean seed fractions together with a summary for conversions to an energy basis.

Seed fraction	Percent	Kcals./g.	Grams carbon per 100g. seed	Grams "work" carbon per 100g. seed
Protein	X_1	4.57	$.532X_1$	$.478(.532X_1)$
Oil	X_2	9.40	$.775X_2$	$.474(.775X_2)$
Residual	X_3†	3.95	$.400X_3$	0

Energy per gram by seed fractions

$E_p/g = 4.57X_1/100$
$E_o/g = 9.40X_2/100$
$E_r/g = 3.95X_3/100$

Energy per acre by seed fractions

$E_p/A = k4.57X_1Y$‡
$E_o/A = k9.40X_2Y$
$(E_p+E_o)/A = k(4.57X_1 + 9.40X_2)Y.$
$E/A = k3.95X_3Y + E_p/A + E_o/A$

† $X_3 = (95-X_1-X_2)$ ‡ Y = YIELD/A

which will permit the comparison of two measurement scales with respect to identifying true genetic differences. The problem is similar to the comparing of different scales of measurement (2, 3, 11, 12) but including genetic theory. The statistic to be developed will be called relative genetic sensitivity.

Consider an arbitrary scale (μ) reflecting the true genetic worth for any character. Then a mapping function, $f(g_1)$, can be found to map the true genotypic scale, (g_1), on μ. With the assumption of an additive genetic model, the mapping function for any character becomes a linear regression. The coefficients $b_{\mu 1}$ for μ on g_1 and $b_{\mu 2}$ for μ on g_2 are defined as scale parameters rather than the normal regression coefficients. The scale parameter can be defined as $b_{\mu 1} = \sigma\mu/\sigma g_1$ (13). Consider that one has a phenotypic observation and wishes to make a statement concerning the true relative genotypic value as measured on the scale μ. The genetic sensitivity (λ_1) would essentially be inversely proportional to the confidence increment in μ, or $\lambda_1 = 1/\Delta t b_{\mu 1}\sigma\epsilon_1$, $\sigma\epsilon^2_1$ being the error variance for g_1. Two measures will have equal genetic sensitivity only if $\Delta t b_{\mu 1}\sigma\epsilon_1 = \Delta t b_{\mu 2}\sigma\epsilon_2$ for any selected level of probability. Relative genetic sensitivity is defined as

$$\phi_{12} = \lambda_1/\lambda_2 = [\sigma g^2_1 \sigma\epsilon^2_2/\sigma g^2_2 \sigma\epsilon^2_1]^{1/2}.$$

With the assumption of independence between λ_1 and λ_2, test procedures have been developed involving statistics proportioned to λ^2_1 and λ^2_2, and the values necessary for tests of significance have been tabulated for cases with limited degrees of freedom (2, 12). For genetic variability studies where the degrees of freedom for progeny and error mean squares are usually large, an approximate standard error for ϕ_{12} can be calculated, using the technique given by Kendall (8) and the asymptotic normal property of statistics estimated from large samples. An approximate standard error for ϕ_{12} is

$$\phi_{12}[(1-\rho^2_1)/n_1 + (1-\rho^2_2)/n_2 + H(2-H)/n_2]^{1/2}/H,$$

where n_1 and n_2 are the degrees of freedom for progeny and for error mean squares, respectively, H is the average heritability for the two characters, $H_1 = r\,\sigma g^2_1/(r\,\sigma g^2_1 + \sigma\epsilon^2_1)$, and ρ_1 and ρ_2 are correlations involving the phenotypes and the error deviates, respectively.[3] The correlations result from the condition that the two phenotypes were measured on the same individual. Hence, λ_1 and λ_2 are not independent. If the data were obtained from two different sources, $\rho_1 = \rho_2 = 0$, and n_1 and n_2 would be the harmonic means of the degrees of freedom for progenies and for error mean squares, respectively.

The definition of relative genetic sensitivity is an attempt to combine the concepts of measurement sensitivity and genetic theory. To have merit, the statistic must not only make biological sense but also be simple to use and easy to visualize. The statistic meets these requirements. If two measures are equally sensitive in identifying genetic differences, $\phi_{12} = 1.0$. One is thus interested in testing the deviation of an estimated ϕ_{12} from 1.0. $\sigma\epsilon^2_1$ is the measure of the reproducibility of a genotype within a population. Although for this study the population is within an environment, both $\sigma\epsilon^2_1$ and H can be defined for environments and involve the genotype by environment interactions. The value ϕ_{12} implies only that two measures have been equally sampled. $100\phi^2_{12}$ would be the number of environments (or replications) needed for measure criterion 2, expressed as a percent of the number of environments (or replications) considered for measure criterion 1, to obtain equal sensitivity in the two measurements for detecting genetic differences.

[3] Two reasonable approximations were made to obtain a usable formula: (i) $V(\sigma g^2_1)/(\sigma g^2_1)^2 \leqq 2(1/n_1 + 1/n_2)/H^2_1$ and (ii) H_1 and H_2 were replaced by H, the average of H_1 and H_2.

RESULTS

The individual year-location analyses as outlined in Table were pooled for the four environments, giving 3,344 legrees of freedom for progeny variability within crosses vithin environments and 3,520 degrees of freedom for a ooled error. There would be no reason to analyze the four nvironments separately unless conflicting results were btained. Actually, the results found for the different enviionments were surprisingly similar. Unless otherwise stated, he analyses will be based upon the pooled analysis.[4]

The primary measures available are yield, percent proein, and percent oil. The protein and oil percentages deterine percent residual. The F ratios and the variance comonents for these four measurement criteria are summarized n Table 4. The ten parents used in the diallel cross system ere selected for their oil potentials and for diversity of enetic backgrounds. The variability found in this study is haracteristic of the F_3 line variability of soybeans. Real enetic variability for residual apparently exists. However, enetic variability for percent protein and oil implies enetic variability for residual due to an artifact created by ie measurement criteria.

Statistical estimation of energy per gram—Consider the ongenetic (error) deviates within a location. A scale rameter of oil to protein would be $\sigma\epsilon_2/\sigma\epsilon_1$. On some ergy measurement scale, the scale parameter between two ed fractions would be 1.0. Thus, $K_{21}\ \sigma\epsilon_2/\sigma\epsilon_1 = 1.0$, and $_{21} = \sigma\epsilon_1/\sigma\epsilon_2$. The error variability affords an estimate of e relative energy scale for two seed fractions. Of equal terest would be the relative energy scale based upon genopic levels. The results are tabulated in Table 5. The 2:1 ange for protein to oil as determined from the experiental errors $(\sigma\epsilon_1/\sigma\epsilon_2)$ is typical of the ratios calculated om other soybean data, and the energy value for oil ould be about twice that for protein, which agrees with e ratio of caloric values for the two fractions. Since the idual is taken as the fraction of seed not accounted for the two fractions estimated chemically, it must be conlered as a reflection of the two fractions and, thus, diffilt to interpret. The energy per gram for residual (3.95 als./g.) could be underestimated. That is, if 4.9 Kcals./g. considered for residual, the ratios based upon error varility and upon caloric values per gram are very similar. e 1.3 to 1 estimate for the change of protein to oil as ermined from the genetic components $(\sigma g_1/\sigma g_2)$ does fit into the caloric value concept. This deviation from 2:1 change on the genetic scale is found in other data. r example, the pooled estimate for the data presented by nson et al. (7) gave 1.6:1. Certainly, an interpretation the energy distribution between the three seed fractions l require more than an energy analysis based on the pective caloric values.

Energy per acre—The transformations required for the asurements of energy per acre in protein, oil, protein s oil, and total are defined in Table 3. The energy-per- measures involve the products of two measures which subject to error or the sum of two or more such prod-. Yield (bu./A.) is the normal measure taken to repret an approximate measure of energy production and will ised as the reference measure. The relative genetic sensities of the energy measures as compared with yield are

[4] Appendix tables giving detailed analyses for the data presented his paper can be obtained upon request from Soybean Investions, Crops Research Division, Beltsville, Md.

TABLE 4—Statistical description of the F_3 soybean line variation within crosses within environments for the primary measures considered.

Statistic	Character			
	Protein (%)	Oil (%)	Residual (%)	Yield (bu./A.)
F ratio	2.08	3.74	1.96	1.62
$\sigma^2_{g_l}$	0.3926	0.2488	0.3017	5.379
$\sigma^2_{\epsilon_l}$	0.7250	0.1816	0.6258	17.243

TABLE 5—Statistical estimates of relative scale changes for the three-soybean seed fractions.

Measurement scale	Ratio		
	O/P	R/P	R/O
Error	2.00	1.08	.54
Genetic	1.26	1.14	.91
Energy	2.06	0.86	.42

TABLE 6—Relative genetic sensitivities with approximate standard errors for the energy measures as compared with yield and the variability in energy associated with yield.

Statistic	Character			
	E_p/A	E_o/A	$(E_p+E_o)/A$	E/A
Relative genetic sensitivity				
ϕ_{ly}	1.01±.037	.99±.037	.98±.035	.99±.034
Relative proportion of variability in E associated with yield				
Phenotypic	0.96	.94	.99	1.00
Genotypic	0.95	.89	.98	.99

given in Table 6. As previously stated, ϕ_{ly} would be a measure of the ability to identify genetic differences on an energy scale as compared with the ability to identify genetic differences on the yield scale. The energy measures are essentially as efficient a selection criterion as yield. The sensitivity ratios are based upon within-environment variability. Actually, both percent oil and percent protein give fairly consistent reactions for environments. If the relative sensitivity ratios were based on the sampling of two or more environments, one might expect the ratios to be greater than 1.0.

The proportions of the phenotypic and the genotypic variability in the energy measures explained by phenotypic and genotypic variability in yield are included in Table 6. The selection for yield is essentially identical with the selection for total energy produced per acre. The results, however, would have been anticipated since the residual fraction is determined as a difference (5% ash assumed) and the seed fractions tend to vary proportionally to energy per gram. One would also conclude that the selection for genetic differences in yield is for all practical purposes the same as selection for the genetic capacity to produce energy protein plus energy oil per acre. This analogy does not follow for the selection for energy protein or energy oil per acre.

Distribution of energy between seed fractions—The genetic correlations between the three seed fractions based on percent by weight and caloric energy per gram must be identical since the two sets of measures differ only by constants. The interpretation of the correlations is difficult since the base units of measure are not equivalent with respect to the energy required to produce the fractions. Therefore, the carbon equivalent unit was defined (Table 3) which expresses each seed fraction in terms of the grams of initial sugar carbon required to produce the respective fractions in a gram of soybean seed. Only the sugar carbon which becomes associated with the seed and the "work" carbons lost in the process of changing the simple sugars to an alternate energy form in the seed are considered. The carbon utilized for growth and metabolism of the plant parts other than seed is not considered.

The total weight of sugar carbon associated with 100 grams of seed would be

$$T_1 = [.532 + (.478)(.532)]X_1 + [.775 + (.474)(.775)]X_2 + .400X_3$$

for an observed seed composition (Table 3). Thus, the proportion of sugar carbon ultimately associated with the protein fraction and the energy required to produce the protein would be

$$p_1 = [.532 + (.478)(.532)]X_1/T_1.$$

Similarly,

$$p_2 = [.775 + (.474)(.775]X_2/T_1, \text{ and}$$
$$p_3 = .400X_3/T_1,$$

where p_2 and p_3 are the proportions of the original hexose carbons associated with the oil and the residual complexes, respectively.

The proportion of the carbons exclusive of the "work" carbons which is associated with protein is $p_1/1.478$ and with oil is $p_2/1.474$. Thus, correlations involving the proportion of sugar carbons associated only with the seed fractions or with the fraction plus "work" carbons are identical.

The average protein and oil based on percent by weight and on the proportionate measures are given in Table 7 for each environment sampled. Environment 15 would be considered an environment resulting in a moderately high level of protein while environment 26 would be considered an environment resulting in extremely high oil. The transformation, thus, does not remove environmental effects. The genetic correlations between protein and oil for the two measurement criteria are also included in Table 7. In general, the level of protein and oil may vary considerably between environments, but the relative performance between genotypes tends to be consistent. The genetic correlation for protein-oil percent reported by Johnson et al. (7) was —.63 for two crosses and two sampled environments. Although the observed correlation (—.54) is based on within-environment variability, it is also based on the variability within 45 crosses.

The phenotypic and genotypic correlations together with the correlations based on error deviates are listed in Table 8 for the percent by weight and the proportionate measures. An increase in the correlations involving protein and oil and a decrease in the correlations involving protein and residual would have been expected on the basis of the modifications in scale introduced by the p_i transformation. However, the magnitude of the change (—.543 to —.852) in the genetic correlation of protein and oil was not anticipated. If the variation in p_1, p_2, and p_3 is proportionate to that found in a multinomial, then the correlations resulting from the restriction, $(p_1 + p_2 + p_3) = 1$, would be —.70, —.39, and —.35, respectively, for P-O, P-R, and O-R. When considering the degrees of freedom available for the variance estimates or the consistency of the correlations in Table 8, one must conclude that the deviation observed for the genetic correlation (—.852) for protein and oil from —.70 (or —.77 based on error deviates) is real. In any event, the genetic controls of the energy distribution into the protein and the oil seed fractions are two very closely interrelated mechanisms. The observation is reasonable since a carbon can be associated with only one of the seed fractions.

The errors for protein and for oil are independent while the residual fraction is estimated by difference. The negative correlation between protein and residual and the positive correlation between oil and residual based upon error deviates would be expected since the laboratory component for percent protein determination is about twice that for oil (unpublished data). Since an independent set of measurements was made for each replication, the genetic correlations involving the residual fraction should not have the restrictions inherent for the correlations based on the error deviates.

Consider the p_i measures relative to the original measure (X_i). The relative genetic sensitivities for each proportionate measure as compared with the corresponding measure by percent weight are given in Table 9. Genetic variability for X_1 and X_2 implies genetic variability for X_3; however p_1 and p_2 can be varied without necessarily varying p_3. Soybean genotypes differ in abilities to produce high energy products, and these differences can be identified with about the same precision as that available for X_3. The p_1 measure is superior to X_1 as a criterion for selecting high protein (relative genetic sensitivity of 1.14). Further, the p_1 measurement scale is measuring modified genetic effects as compared with X_1 measurement scale as indicated by the proportion of the genetic variability in p_1 explained by X_1. In the redistribution of genetic variability through the p_i transformation, the p_1 measure apparently contains genetic effects formerly associated with the residual (and possibly error) as measured on the X_i scale.

Finally, with the p_i transformation one can test indirectly for the association of the genetic mechanisms involved in

TABLE 7—Average protein and oil levels and genetic correlations between protein and oil within environments as measured on the percent (X_i) and proportionate (p_i) scales.

Location-year	Protein		Oil		Genetic correlation
	X_1 (%)	p_1	X_2 (%)	p_2	X_1X_2 p_1p_2
15	42.8	.478	21.3	.346	-.421 -.8
25	41.9	.469	21.7	.352	-.525 -.8
16	41.1	.459	22.3	.362	-.407 -.8
26	40.5	.447	23.8	.381	-.759 -.9
Pooled	41.6	.463	22.2	.360	-.543 -.8

TABLE 8—Phenotypic, genotypic, and error correlations for percent by weight (X_i) and the proportionate (p_i) scales based upon a pooled analysis.

Variables	Correlations		
	Phenotypic	Genotypic	Error
Protein-Oil			
Percent (X_i)	-.474	-.543	-.
Proportionate (p_i)	-.816	-.852	-.
Protein-Residual			
Percent (X_i)	-.756	-.647	-.
Proportionate (p_i)	-.465	-.289	-.
Oil-Residual			
Percent (X_i)	-.218	-.288	-.
Proportionate (p_i)	-.132	-.255	+.

TABLE 9—Relative genetic sensitivities with approximate standard error of the proportionate measure (p_i) as compared with the percent measure (X_i) and the variability of the p_i associated with X_i.

Statistic	Character		
	Protein	Oil	Residual
Relative genetic sensitivity $\phi\ p_iX_i$	1.14±.034	.97±.023	1.02±.03
Relative proportion of variability in p_i associated with X_i			
Phenotypic	.924	.966	
Genotypic	.914	.978	

TABLE 10—Phenotypic, genotypic, and error correlations involving the ratio p_1/p_2 and the proportionate residual (p_3) estimated for each environment.

Environment	Correlations		
	Phenotypic	Genotypic	Error
15	-.122	+.208	-.51
25	-.132	-.037	-.40
16	-.380	-.011	-.75
26	-.466	-.499	-.46
Pooled	-.253	-.054	-.54

the synthesis of protein (or oil) and the synthesis of residual. The correlations between the ratio p_1/p_2 and p_3 as measured at the phenotypic, genotypic, and error levels are given in Table 10. The data are summarized by environments. Since the residual is determined as a difference, the negative correlations involving error could reflect the differential variability of the protein and the oil determinations. The average genetic correlation of approximately zero indicates that on the average the sugar carbons are distributed to protein and oil in the same proportions in genotypes with a high residual as in genotypes with a low residual. However, this genotypic expression apparently is not a consistent feature for the environments sampled. For the two environments (25 and 16) which have protein and oil levels characteristic for the populations studied, the genetic correlations are essentially zero; however, for the environment with the moderately high protein (15) the genetic correlation is positive while for the environment with the extremely high oil level (26) the genetic correlation is negative. On the surface these correlations appear contradictory. The ramifications for these correlations will be considered in the Discussion.

DISCUSSION

The ideal measurement scale would be one which directly reflects the primary gene effects. Total seed energy reflects the ability of a genotype to synthesize energy. A soybean genotype apparently has the capacity to balance growth progressively with its yield potential by continued podding or flower dropping. At maturity the seed energy is stored with a seed size and an energy distribution between seed fractions characteristic of the genotype. Therefore, energy produced by a genotype as measured by some function of yield is considered to be a more realistic measure of gene effects than any component approach involving seed size, seeds per fruiting body, fruiting bodies per plant, etc., which involve measures incidental to the basic genetic problem. Grafius' statement (5) "Yield is thus an artifact" is a meaningless statement referenced to a geometric interpretation which is incidental to the basic genetic problem.

The assumption has been made that the composition of fatty acids in soybean oil and of amino acids in soybean protein is constant over genotypes, or that the description for the average seed composition approximates specific fractions within practical limits. With more complete biochemical information, a more exact p_i transformation could be defined. A realistic and acceptable transformation must have been developed to obtain the correlations reported in this study.

Distribution of sugar carbons between seed fractions—The genetic capacity of a plant to produce total seed energy is limited by its genetic capacity to produce sugar energy. With the carbon equivalent concept one expresses each seed fraction in terms of an equivalent base measure, the grams of sugar carbon required to produce a fraction including "work" carbon. Energy associated with plant growth and metabolism is not considered. Let us start with the premise that the initial sugar carbons are not predestined or are not associated with a particular ultimate energy fraction. That is, a genetic potential to produce sugar energy can be visualized.

The residual comprises a structural fraction (primarily cellulose) and a soluble carbohydrate fraction. The structural fraction would be considered essentially a constant feature, at least for the improved soybean genotypes. The soluble carbohydrate fraction would represent a reservoir. Relative to the genetic variability found in the protein and oil fractions, the genetic variability for the residual is restrictive (Table 8). With complete restriction in genetic variability for residual the genetic correlation between the proportions of sugar carbon being associated with the protein and oil complexes (p_1 and p_2) would be -1.0. The high genetic correlation between p_1 and p_2 is a direct expression of the restricted genetic variability in the residual fraction. Similar information is available from an examination of the genetic components of variation. The relative genetic variances for the p_1, p_2, and p_3 measurement scales are 1.00, .98, and .29, respectively, as compared with 1.00, .93, and .58, respectively, for the multinomial.

Apparently, through selection, genotypes with a minimum cellulose structure have been evolved. Although selection can be made for the efficiency of a genotype to produce high energy forms, the progress would be restricted. Thus, sugar carbons tend to become associated with one of the two pathways of synthesis for protein or oil, and the protein-oil synthesis could be considered as a competitive balance for the sugar carbons between the genes for protein and for oil. Hence, a population may be genetically homozygous for protein factors but segregating for oil factors, yet the relative genetic variability for both protein and oil would be similar to a population segregating for both protein and oil genes. Since genetic variability exists for oil within the high oil lines, one can select for high protein from high oil lines. However, as one eliminates the factors for high oil, he probably will be faced with inefficient genotypes with respect to conversion of residual to high energy forms.

On the surface the signs of the genetic correlations as given in Table 10 appear to be reversed for locations 15 and 26. The postulation has been made that the protein and the oil synthesis can be conceived as an association of sugar carbons with alternative energy forms and a competitive balance for the sugar carbons between the genetic structure for protein and for oil. In an environment conducive to a high protein level, environmental factors could favor protein synthesis and tip the competitive balance in favor of protein factors. The synthesis for oil in the high oil genotypes would draw the energy in part from the soluble carbohydrates in the residual giving a positive genetic correlation between p_1/p_2 and p_3. In environments conducive to a high oil level the converse would follow giving a negative genetic correlation between p_1/p_2 and p_3.

Relative scale change between seed fractions—Attention has been given (Table 5) to the relative scale change for the seed fractions and the similarity of the estimated scales to the caloric values for each seed fraction. However, discrepancies were noted when the measurements were considered at the genotypic level. Let us re-examine the data in light of the concept of carbon equivalents. The weight of sugar carbon used to produce a gram of soybean seed is $[(.7863)X_1 + (1.1424)X_2 + (.4)(95-X_1-X_2)]/100$ (Table 3). With a constant genotypic capacity to produce sugar, a change in oil (ΔX_2) reflects directly the change in protein, $\Delta X_1 = -.7424\,\Delta X_2/.3863$. Therefore, the scale change of protein to oil on the error level would be $\Delta X_1/\Delta X_2 = -1.92$, approximately. The scale change of protein to oil between genotypes would involve different genetic capacities to produce energy and relative genetic

variability for the residual fraction. With no genetic variability in residual (p_3), then $\Delta p_2 = -\Delta p_1$, and $\Delta X_1/\Delta X_2 = -1.1424/.7863 = -1.45$ as measured on the X_1 scale. Thus, in the selected high oil lines one would expect the scale change of perhaps 1.5–1.6 for X_1 to X_2 as measured on the genotypic scale. The breeder will probably find that a 2% gain in protein will not be realized for each 1% loss in oil when converting to high protein lines.

However, with the assumption that the change in protein to oil percentages averages 1.92:1, the gram carbon of sugars required to produce one gram of soybean seed (Table 3) is essentially constant for a normal range of protein and oil percentages. For an average protein to oil change of less than 1.92:1, the percent protein and grams of seed produced for a constant amount of sugar energy are positively correlated. Since approximately .40, .79, and 1.14 grams of sugar carbon are required to produce a gram of soybean residual, protein, and oil, respectively, increased protein or residual at the expense of oil implies higher yield for a constant energy potential.

The negative correlations reported by Johnson et al. (7) between yield and percent protein could simply be a manifestation of these results. Within an environment conducive to high oil (say limiting nitrogen supply) the high protein genotypes tend to be associated with a low residual. A reduced residual implies a reduced yield for a constant initial energy. These results were also observed in the present study where the genetic correlations between yield and percent protein were estimated to be +.142, +.002, −.060, and −.203 for environments 15, 25, 16, and 26, respectively. These correlations are to be compared with the correlations in Table 10.

Ramifications for selection—The transformations considered in this paper were made so that the measurement scale would represent more directly gene effects and an interpretation could be made of the genetic mechanisms involving the production of the protein, oil, and residual fractions. In a practical breeding program for high protein the measure percent protein (X_1) still has merit as a selection criterion. Selection pressures on percent protein would exert a greater pressure for reducing the residual fraction than selection pressures on the proportionate (p_1) scale (Table 8), principally due to the scale parameters for the two measures. On the other hand, if one were to develop a feed bean high in protein under an economic structure where the oil had a low economic value, the proportionate measurement scale would be recommended since p_1 more directly measures the genetic mechanisms for protein synthesis, the p_1 measure has a significantly greater relative genetic sensitivity than percent protein, and the residual has food value for feed purposes. The combination must be associated with high yield unless there are physiological restrictions, such as inadequate nitrogen during the critical period of conversion of protein precursors to seed protein.

SUMMARY

The caloric values and the weight of initial sugar carbons required to produce a fraction were developed for each of the soybean seed fractions. The caloric measurement scale led to agronomic evaluations but was considered inadequate for a complete genetic analysis. The latter measure permitted a transformation (p_1) expressing the proportion of the initial sugar carbons associated with each seed fraction. With the assumption that a genotype has unique capacity to produce initial sugar energy, interpretable concepts of protein, oil, and residual synthesis resulted.

A statistic, relative genetic sensitivity, was developed to compare two measurement scales as selection criteria.

Energy per acre based on the seed fractions or upon total energy was essentially as efficient a selection criterion as yield. The selection for high yield was essentially identical genetically with the selection for total energy per acre.

Genotypes differ in the ability to convert the residual fraction to oil or protein. Within highly selected lines, however, the genetic variability in the residual was found to be restricted as compared with the genetic variability for oil and protein. Thus, the genetic correlation between protein and oil was high (−.852) as measured on the proportionate (p_1) scale.

The scale change for protein to oil percentages was calculated as 1.92 for random deviates and between 1.45 and 1.92 for changes on the genetic scale, based on the concept of carbon equivalents. The gain in percent protein for each loss of 1% oil in a breeding program for high protein would be expected to be less than 2%, perhaps in the range of 1.5–1.6%, but for a constant energy potential yield should be increased. Experimental estimates for the scale change were comparable to the theoretical values.

The average genetic correlation between percent protein and yield estimated from the data was negative but essentially zero, although this may vary with environments.

One would expect the genetic correlations between percent protein and yield to be small and positive. Since approximately .79, 1.14, and .40 grams of sugar carbon were required to produce one gram of protein, oil, and residual, respectively, there should be no restrictions in producing a high yielding, low oil bean and, if nitrogen is not limiting or other physiological restrictions are not present, a high yielding, high protein bean.

LITERATURE CITED

1. Block, R. J., and Weiss, K. W. Amino Acid Handbook: Methods and results of protein analysis. C. C. Thomas, New York. 1956.
2. Bradley, R. A., and Schumann, D. E. W. The comparison of the sensitivities of similar experiments: Applications. Biometrics 13:496–510. 1957.
3. Cochran, W. G. The comparison of different scales of measurement for experimental results. Annals Math. Stat. 14:205–216. 1943.
4. Fruton, J. S., and Simmonds, S. General Biochemistry, Chapter 32. John Wiley and Sons Inc., New York. 2nd Ed. 1958.
5. Grafius, J. E. Heterosis in barley. Agron. J. 51:551–554. 1959.
6. Howell, Robert W. Physiology of the soybean. Advances in Agronomy, 12. Academic Press, N. Y. 1960.
7. Johnson, Herbert W., Robinson, H. F., and Comstock, R. E. Genotypic and phenotypic correlations in soybeans and their implications in selection. Agron. J. 47:477–483. 1955.
8. Kendall, M. G. The advanced theory of statistics, Vol. 1, Chapter 9. Hafner Pub. Co., New York. 5th Ed. 1952.
9. Kharasch, M. S. Heats of combustion of organic compounds. Bur. Stds. J. Res. 2:359–430. 1929.
10. Leffel, R. C., and Hanson, W. D. Early generation testing of diallel crosses of soybeans. (Submitted for publication. Crops Sci., 1960.)
11. Mandel, J., and Stiekler, R. D. Sensitivity—a criterion for the comparison of methods of tests. Bur. Stds. J. Res. 53:155–159. 1954.
12. Schumann, D. E. W., and Bradley, R. A. The comparison of the sensitivities of similar experiments: Model II of the analysis of variance. Biometrics 15:405–416. 1959.
13. Teissier, Georges. La relation d'allométrie: sa signification statistique et biologique. Biometrics 4:14–53. 1948.

CROP SCIENCE

Volume 24 January–February 1984 Number 1

Selection Indexes to Modify Protein Concentration of Soybean Seeds[1]

S. J. Openshaw and H. H. Hadley[2]

ABSTRACT

The purpose of this study was to investigate the effectiveness of selection based upon selection indexes to modify protein concentration of soybean [*Glycine max* (L.) Merr.] seeds. Selection criteria were compared by evaluating F_4 performance of F_2-derived lines identified in the F_3 as being in the upper 10% of the population. Smith-Hazel indexes designed to maximize the gain in protein by including oil or oil and sugar as secondary traits were not more effective than direct selection, which resulted in changes of 1.42% protein and −1.14% oil. Smith-Hazel indexes which gave protein and oil equal economic weights, either including or not including sugar as a secondary trait, and a desired gains index with desired gains set at 1.0% protein and 0.25% oil were all similar in results to selection based on the sum of protein + oil. Among these methods, changes ranged from 0.15 to 0.31% for protein and from −0.01 to 0.19% for oil. An index designed to change protein by 1.25% and oil by −0.75% produced observed changes of 1.31 and −0.81%. An index designed to increase protein as much as possible while holding oil constant resulted in changes of 0.45% protein and −0.13% oil. This index was more effective than selection for protein following culling oil at the mean, which resulted in little change for either trait.

Additional index words: Glycine max (L.) Merr., Oil, Sugar.

Soybean [*Glycine max* (L.) Merr.] occupies an important place in the world market because of its oil and protein content. Typically, soybean cultivars grown in the the USA contain about 41% protein, 21% oil, and 11% soluble carbohydrates. Both oil and protein are highly heritable. Based upon information from several soybean researchers, Johnson and Bernard (1963) presented estimates of expected heritabilities of 0.63 for protein and 0.67 for oil when selection was based on means of two replications of F_3 or later generation lines grown at two locations. They estimated the expected genetic correlation between protein and oil to be −0.60. Thus, selection to increase either oil or protein would be expected to be effective but would be expected to lead to a decrease in the other. Although the contribution of protein to the value of soybeans has increased, at present it would not be practical to increase protein at the expense of oil. Thorne and Fehr (1970) proposed that selection for the sum of protein + oil might increase protein without a marked decrease in oil. This amounts to selection based on an index in which each trait is given equal weight.

Smith (1936) proposed an index of the form $I = \Sigma b_i X_i$ where b_i, the weight given to trait X_i, is derived such that the index should best predict the true worth of a line. Hazel (1943) extended the concept by considering only the additive effects of genes in regard to the genetic variances and covariances used in the calculations. Hazel and Lush (1942) and Young (1961) showed that selection based on such an index was theoretically never less efficient than selection based on independent culling levels or than the use of tandem selection for each trait.

Selection based on the Smith-Hazel type of index leads to changes in the breeding values of individual traits. Kempthorne and Nordskog (1959) developed a restricted selection index for situations in which it is desired to increase the aggregate genetic value as much as possible while holding the values of some traits constant.

Both of the above indexes require the assigning of relative economic weights to each trait. This enables secondary traits to be used to increase the progress made with primary traits (by assigning economic values of zero to the secondary traits) but limits the usefulness of such indexes to those cases where the breeder is able to designate appropriate economic values. Pesek and Baker (1969) presented a desired gains index which avoids the need for economic weights provided the breeder specifies the genetic gain desired for each trait.

A major handicap to the effective use of all of the above indexes is that the variances and covariances needed for their construction are subject to large sampling errors. Brim et al. (1959) suggested that an index which weighted each character only by its relative worth would circumvent the problem, but they did not speculate on the effectiveness of such an index. Selection based on the sum of oil + protein would be such a base index (Williams, 1962), in which each trait is assigned equal economic value.

The purpose of this study was to investigate the effectiveness of selection based on indexes designed to increase the protein concentration of soybean. Genetic correlations between sugar and protein were reported as −0.68 and −0.60 in two populations studied by Openshaw and Hadley (1981). Thus, we used sugar and oil as secondary traits to attempt to increase the gain in protein resulting from selection.

[1] Contribution from the Dep. of Agronomy, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801. Supported by funds from the Illinois Agric. Exp. Stn. Received 17 May 1982.

[2] Former graduate research assistant (present address: DEKALB-PFIZER GENETICS, 3100 Sycamore Road, DeKalb, IL 60115) and professor of plant genetics, respectively.

Restricted and desired gains indexes and a base index composed of the sum of oil + protein were compared for their effectiveness in increasing protein while controlling the corresponding change in oil.

MATERIALS AND METHODS

Two populations were used in this study. Population I was derived from the cross HHP × 'Mandell' and Population II from 'Protana' × 'Williams'. HHP is a high-protein experimental line developed by H. H. Hadley by introgressing germplasm from *Glycine soja* Sieb. and Zucc. into that of *G. max*. All experiments were grown on the Agronomy South Farm of the Univ. of Illinois, Urbana, Ill. About 200 F_2 plants in 1976, 150 F_3 lines in 1977, and 150 F_4 lines in 1978 were grown from Population I. About 500 F_2 plants in 1977 and 300 F_3 lines in 1978 were grown from Population II.

F_2 plants were grown in single-plant hills spaced 30 × 38 cm apart. One hundred fifty F_2 plants from Population I and 300 from Population II were chosen on the basis of sufficient seed. F_3 progenies were planted in single-row plots 2.5 m long, spaced 0.76 m apart, and were seeded at a rate of 40 seeds per row. Populations were divided into blocks consisting of 30 F_3 lines each and planted as a blocks-in-replications design with two replications. Each block also contained both parental lines. For Population I, an F_2-derived line in the F_4 was grown from bulked seed from F_3 plants of each F_2-derived family. These lines were planted in the same experimental design as were F_3's, with lines being rerandomized within blocks. In F_3 and F_4 lines, the middle 1.8 m of row was harvested and seeds were analyzed for protein, oil, and sugar percentage.

Values for protein and oil were obtained using a near-infrared reflectance grain analyzer (Hymowitz et al., 1974). Values for sugar were obtained using a phenol-sulfuric acid method described by Openshaw and Hadley (1981).

Estimates of variances and covariances were obtained by equating observed mean-squares and cross-products from the analysis of variance or covariance with their expectations and solving for desired variance (covariance) components. The genetic variance among F_2-derived lines is composed of the additive plus one-fourth of the dominance variance in the F_3 and of the additive plus 1/16 of the dominance variance in the F_4. Studies have shown that additive genetic variance is the predominant component of genetic variance for characters of economic importance in soybeans (Brim and Cockerham, 1961; Hanson et al., 1967; Croissant and Torrie, 1971).

Index weights for the Smith-Hazel indexes designed to maximize gain were calculated by solving the equations $b = P^{-1}Ga$ where P^{-1} is the inverse of the phenotypic variance-covariance matrix, G is the genotypic variance-covariance matrix, and a is the vector of relative economic values. In this study, each primary trait was assigned a value of 1 and each secondary trait a value of 0.

Index weights designed to achieve desired gains were calculated by solving the equations $b = G^{-1}h$ where G^{-1} is the inverse of the genotypic variance-covariance matrix and h is the vector of desired gains (Pesek and Baker, 1969). Index weights designed to improve protein as much as possible while holding oil constant were calculated from the above equation by setting the desired gain of oil at 0. Relative index weights so obtained are equivalent to those obtained with the formula given by Kempthorne and Nordskog (1959).

Predicted gains for individual traits were calculated for the Smith-Hazel index as $G_i = \Sigma b_j G_{ij} k/\sigma_I$ where G_i is the predicted gain for the i*th* trait, b_j is the index weight of the j*th* trait, G_{ij} is the genotypic covariance of the i*th* and j*th* trait, k is the standardized selection differential (1.74 for Population I and 1.76 for Population II), and $\sigma^2_I = b'Pb$ where b' is the transpose of the vector of index weights.

Predicted gains resulting from direct selection based on two replications at one location were calculated as $G = h^2\sigma_p k$ where h^2 is the heritability of the trait under selection as estimated from variance components derived from the analysis of variance, σ_p is the phenotypic standard deviation, and k is the standardized selection differential. Predicted correlated responses in trait j resulting from selection for trait i were calculated as $G_j = (h_i^2 h_j^2)^{1/2} r_{G_{ij}} \sigma_{P_j} k$ where h_i^2 and h_j^2 are heritability estimates for the i*th* and j*th* trait, $r_{G_{ij}}$ is the genotypic correlation between traits i and j, estimated from the analysis of covariance, σ_{P_j} is the phenotypic standard deviation of the j*th* trait, and k is the standardized selection differential. Predicted responses were adjusted for differences in phenotypic variance between years.

In all selection experiments, the upper three lines were identified in each block for a proportion selected of 10%. Response to selection was calculated as the difference of the mean values for the progenies of selected lines from the population mean.

RESULTS AND DISCUSSION

Differences between the parents of Population I for percentage protein and oil were considerably greater than those between the parents of Population II while differences in percentage sugar seemed similar for the two sets of parents (Table 1). These parents were chosen with the thought that heritabilities for protein and for oil would be high in Population I and intermediate or low in Population II.

Estimates of heritability and phenotypic and genotypic correlations were calculated from the analysis of variance and covariance of F_3 generations (Table 2). Because the trials were grown in only one environment, these estimates could be inflated by genotype × environment interactions, which are frequently not important for these traits. As expected, genetic variances were greater in Population I than in Population II for protein (1.47 vs. 0.33), oil (1.20 vs. 0.35) and the sum of protein + oil (0.47 vs. 0.22). As a result, heritabilities for these traits were greater in Population I although they were also high in Population II. Heritabilities for sugar were similar in the two populations.

Genetic correlations were similar to phenotypic correlations in both populations. Protein was nega-

Table 1. Chemical composition of seeds of parental genotypes.†

Parental genotype	Year grown	Protein	Oil	Sugar
			%	
HHP	1977	49.1	13.9	7.7
Mandell	1977	46.4	18.7	8.5
Standard error		0.34	0.24	0.17
HHP	1978	49.6	13.8	7.9
Mandell	1978	46.2	19.1	8.4
Standard error		0.29	0.29	0.20
Protana	1978	45.9	19.6	9.0
Williams	1978	45.2	20.8	8.6
Standard error		0.27	0.31	0.17

† Mean of three replications of plots.

tively correlated with oil and with sugar, being most strongly so in Population I. The positive correlation between oil and sugar was also strongest in Population I. In populations studied by Thorne and Fehr (1970) and Miller and Fehr (1979) correlations of protein + oil with protein were positive and higher than those found in this study while those with oil were very low and tended to be negative. The correlation of protein + oil with oil was low and positive in Population I and was slightly larger than that with protein in Population II.

The effectiveness of selecting for the sum of protein + oil + sugar was investigated in order to see if the proportion of the seed containing these major storage nutrients could be increased. Genetic variances for the sum of the three mentioned components were estimated in the F_3 as 0.38 for Population I and 0.19 for Population II with corresponding heritabilities of 0.75 ± 0.04 and 0.59 ± 0.05. Using the sum of protein + oil + sugar as a selection criteria in the F_3 of Population I resulted in a gain of 0.48 percentage points in the F_4 (significantly different from zero at the 0.01 level). This response was composed of a decrease of 0.13% protein (significantly less than its predicted value at the 0.01 level) and increases of 0.47% oil and 0.14% sugar (both not significantly different from predicted values at the 0.05 level). These results show that it is possible to increase the proportion of the seed consisting of the three major storage nutrients.

Three selection criteria designed to increase percentage protein were compared for effectiveness in the F_3: 1) direct selection for percentage protein, 2) Index-1 which included protein as the primary trait and oil as a secondary trait, and 3) Index-2 which included protein as the primary trait and both oil and sugar as secondary traits (Table 3). In formulating the indexes, the economic weights of secondary traits were given values of zero. These indexes were designed to take advantage of the genetic correlations of secondary traits with protein in order to maximize the gain for protein. Predicted responses in both populations indicate no practical advantages of the indexes over direct selection, although observed responses in the F_4 of Population I were slightly greater for the index methods. In Population I both indexes chose the same 15 lines and 12 of the 15 lines selected by direct selection. The observed decrease in oil accompanying selection based on the indexes was −1.30 as compared to −1.14 when based on protein alone. This trend might be expected since the indexes select directly against oil. Relative index weights for oil were considerably lower in Population II and direct selection there chose the same 30 lines as did Index-1 and 29 of those chosen by Index-2.

Increases in percentage protein resulting from direct selection or from an index designed to maximize gain in protein were accompanied by rather substantial observed decreases in oil. Several criteria were applied to the F_3 of Population I to compare them for their effectiveness in increasing protein while reducing or eliminating the accompanying decrease in oil (Table 4): 1) the sum of percentage protein + oil, as proposed by Thorne and Fehr (1970), 2) Index-3, a Smith-Hazel index which included protein and oil given equal economic weights, 3) Index-4, a Smith-Hazel index which included protein and oil given equal economic weights and sugar given an economic weight of zero, 4) Index-5, a desired gains index in which desired gains were set at 1.0% protein and 0.25% oil, 5) Index-6, a desired gains index in which desired gains were set at 1.25% protein and −0.75% oil, 6) Index-7, in which desired gains were set to increase protein as much as possible while holding the change in oil at zero, and 7) percentage protein among those lines that were greater than or equal to the population mean for percent oil.

Table 2. F_3 estimates of phenotypic correlations† (above the diagonal), genotypic correlations (below the diagonal with standard errors), and heritability† (on the diagonal with standard errors).

	Protein	Oil	Protein + oil	Sugar
Population I				
Protein	0.90 ± 0.02	−0.80**	0.48**	−0.59**
Oil	−0.83 ± 0.03	0.93 ± 0.01	0.14	0.40**
Protein + oil	0.44 ± 0.08	0.13 ± 0.09	0.78 ± 0.04	−0.39**
Sugar	−0.68 ± 0.07	0.49 ± 0.08	−0.42 ± 0.10	0.72 ± 0.05
Population II				
Protein	0.75 ± 0.03	−0.65**	0.36**	−0.50**
Oil	−0.68 ± 0.05	0.71 ± 0.03	0.47**	0.21**
Protein + oil	0.37 ± 0.08	0.43 ± 0.07	0.67 ± 0.04	−0.32**
Sugar	−0.60 ± 0.06	0.26 ± 0.08	−0.41 ± 0.08	0.67 ± 0.04

** Phenotypic correlation significant at the 0.01 probability level.
† Based on two replications.

Table 3. Comparison of selection criteria in F_3 designed to maximize gain in percentage protein.

Selection criterion	Relative index weight			F_4 response	
	Protein %	Oil %	Sugar %	Predicted	Observed
				% Protein	
Population I					
Protein	1.0	0	0	1.64	1. 2
Index-1	1.0	−0.154	0	1.65	1.64
Index-2	1.0	−0.175	−0.184	1.65	1.64
Population II					
Protein	1.0	0	0	0.88	–
Index-1	1.0	−0.012	0	0.88	–
Index-2	1.0	−0.031	−0.172	0.88	–
Standard error					0.09

Table 4. Comparison of selected criteria in F_3 of Population I designed to increase protein while controlling change in oil.

Selection criterion	Relative index weight			F_4 response			
				Predicted		Observed	
	Protein %	Oil %	Sugar %	Protein	Oil	Protein	Oil
				%			
Protein	1.0	0	0	1.64	−1.32	1.42	−1.14
Protein + oil	1.0	1.0	0	0.68	0.20	0.15	0.19
Index-3	1.0	1.002	0	0.67	0.21	0.23	0.15
Index-4	1.0	1.009	−0.041	0.67	0.21	0.31	−0.01
Index-7	1.0	0.916	0	–	0	0.45	−0.13
				Desired		Observed	
Index-5	1.0	0.994	0	1.0	0.25	0.28	0.14
Index-6	1.0	0.409	0	1.25	−0.75	1.31	−0.81
Cull oil at mean	–	–	–	–	0	0.03	0.11
Standard error						0.09	0.09

The selection indexes designed to maximize the gain in combined percentage protein and oil (Indexes -3 and -4) and that in which desired gains were positive for the two traits (Index-5) were virtually identical to selection for the sum of protein + oil. Theoretical expectations were much the same for these methods although there was some variation among observed results. Gains in percentage oil were usually comparable to predicted values but gains in protein fell far short of expectation.

Indexes -3 and -4 were designed to maximize the gain in economic worth which, in this case, is the sum of gains in percentage oil and percentage protein since both traits are assigned equal economic values. Predicted gains in economic worth for the two index methods and for selection for the sum of protein + oil were virtually identical. Observed gains were significantly different from zero (0.05 level) but fell short of predicted values due to the lack of response in protein mentioned above. Observed gains in economic worth ranged from 0.30 to 0.38, there being no significant difference among responses (0.05 level).

Desired gains for Index-6, set at 1.25% protein and −0.75% oil, showed good agreement with the corresponding observed changes of 1.31 and −0.81%. Index-7, designed to hold oil constant while increasing protein, seemed reasonably effective with an increase of 0.45% protein and a decrase of 0.13% oil. Although the achieved increase was only 32% of that resulting from selection for protein alone, the accompanying decrease in oil was reduced to only 11% of that which accompanied the direct selection. Index-7 seemed superior to selection for protein after culling for oil at the population mean, which resulted in little change for either trait.

In summary, indexes using oil and sugar as secondary traits designed to maximize gain in percentage protein were not superior to direct selection for protein. Methods designed to increase both protein and oil showed limited success, although it appeared that slow progress in that direction could be made. The index methods using desired gains which attempted to hold oil constant or control its decrease while increasing protein seemed quite effective and were more effective than selection for protein following culling oil at the mean.

Although the desired gains indexes were effective in controlling changes in both oil and protein concentration, identification of the best line from among a group of homozygous or near-homozygous lines as the end-product of selection could be done just as effectively by direct observation since heritabilities are high for both traits. However, a desired gains index could be of value in a recurrent selection program in which it is desired to identify a group of genotypes for recombination.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Dr. C. Chaven in developing the laboratory procedures used and of C. E. Brokoski and E. M. Miller for their help in various aspects of the study.

REFERENCES

1. Brim, C.A., and C.C. Cockerham. 1961. Inheritance of quantitative characters in soybeans. Crop Sci. 1:187–190.
2. ——, H.W. Johnson, and C.C. Cockerham. 1959. Multiple selection criteria in soybeans. Agron. J. 51:42–46.
3. Croissant, G.L., and J.H. Torrie. 1971. Evidence of nonadditive effects and linkage in two hybrid populations of soybeans. Crop Sci. 11:675–677.
4. Hanson, W.D., A.H. Probst, and B.E. Caldwell. 1967. Evaluation of a population of soybean genotypes with implications for improving self-pollinated crops. Crop Sci. 7:99–103.
5. Hazel, L.N. 1943. The genetic basis for constructing selection indexes. Genetics 28:476–490.
6. ——, and J.L. Lush. 1942. The efficiency of three methods of selection. J. Hered. 33:393–399.
7. Hymowitz, T., J.W. Dudley, F.I. Collins, and C.M. Brown. 1974. Estimations of protein and oil concentration in corn, soybean, and oat seed by near-infrared light reflectance. Crop Sci. 14:713–715.
8. Johnson, H.W., and R.L. Bernard. 1963. Soybean genetics and breeding. p. 1–73. *In* A.G. Norman (ed.) The soybean. Academic Press, New York.
9. Kempthorne, O., and A.W. Nordskog. 1959. Restricted selection indexes. Biometrics 15:10–19.
10. Miller, J.E., and W.R. Fehr. 1979. Direct and indirect recurrent selection for protein in soybeans. Crop Sci. 19:101–106.
11. Openshaw, S.J., and H.H. Hadley. 1981. Selection to modify sugar content of soybean seeds. Crop Sci. 21:805–808.
12. Pesek, J., and R.J. Baker. 1969. Desired improvement in relation to selection indices. Can. J. Plant Sci. 49:803–804.
13. Smith, H.F. 1936. A discriminant function for plant selection. Ann. Eugen. 7:240–250.
14. Thorne, J.C., and W.R. Fehr. 1970. Incorporation of high-protein exotic germplasm into soybean populations by 2- and 3-way crosses. Crop Sci. 10:652–655.
15. Williams, J.S. 1962. The evaluation of a selection index. Biometrics 18:375–393.
16. Young, S.S.Y. 1961. A further examination of the relative efficiency of three methods of selection for genetic gains under less restricted conditions. Genet. Res. 2:106–121.